AFOSR - TR - 72 - 1547

ARO-D Report 71-3

AD 747029

PROCEEDINGS OF THE SIXTEENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

MINIMUM DISCRIMINATION INFORMATION ESTIMATION AND APPLICATION

Approved for public release; distribution unlimited.
The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Approved for public release;
distribution unlimited.

Sponsored by
The Army Mathematics Steering Committee
on Behalf of

41

THE OFFICE OF THE CHIEF OF RESEARCH AND DEVELOPMENT

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) The George Washington University Statistics Department Washington, D. C. 20006	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

MINIMUM DISCRIMINATION INFORMATION ESTIMATION AND APPLICATION

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)

Solomon Kullback

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March 1972	38	17

8a. CONTRACT OR GRANT NO AFOSR 68-1513 b. PROJECT NO. 9749 c. 61102F d. 681304	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR - TR - 72 - 1542

10. DISTRIBUTION STATEMENT

A. Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES PROCEEDINGS Proceedings of the 16th Conference on the Design of Experiments in Army Research Development and Testing, Fort Lee, Va 21-23 Oct 1970.	12. SPONSORING MILITARY ACTIVITY Air Force Office of Scientific Research/NM 1400 Wilson Blvd Arlington, Virginia 22209

13. ABSTRACT

~~This~~ paper presents in some detail the application of the minimum discrimination information theorem to the analysis of multidimensional contingency tables. It is shown that the form of the minimum discrimination information estimate as a member of an exponential family provides a regression expression for the logarithm of the estimate. Computational procedures for the evaluation of the regression parameters and the minimum discrimination information estimates are described along with the tests for the hypotheses as provided by the minimum discrimination information statistics.

1

DD FORM 1 NOV 65 1473

Security Classification

Security Classification

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Contingency Table Analysis						
Minimum Discrimination Information Estimation						
Log-linear representation						
Tests for hypotheses						

Security Classification

Minimum Discrimination Information Estimation and Application

S. Kullback
The George Washington University
Washington, D. C. 20006

Abstract

This paper presents in some detail the application of the minimum discrimination information theorem to the analysis of multidimensional contingency tables. It is shown that the form of the minimum discrimination information estimate as a member of an exponential family provides a regression expression for the logarithm of the estimate. Computational procedures for the evaluation of the regression parameters and the minimum discrimination information estimates are described along with the tests for the hypotheses as provided by the minimum discrimination information statistics.

0. Introduction. This paper is related to [9] and [10] in which certain basic techniques and procedures were presented for the

[1] Supported in part by the Air Force Office of Scientific Research, ~~Office of Aerospace Research,~~ United States Air Force, under Grant AFOSR-68-1513.

analysis of multidimensional contingency tables. In this paper we shall examine the underlying theory in greater detail and present one important area of application. In particular we shall detail the close analogy of this application with multivariate regression analysis. Although the ingredients of the underlying theory were discussed in [11] it seems necessary and desirable to present these ideas here in greater detail. We also remark that a more extensive computer program than that described in [6] and [9] has been prepared by Professor Ireland of The George Washington University. This new program can handle tables of higher dimension than four-way contingency tables and also provides the values of additional useful parameters.

It should be pointed out that there are other areas of application of minimum discrimination information estimation than that considered in detail in this paper, for example, [3], [4], [5], [7], [11], [12], [13], [14]. The particular application we shall consider here can be described as fitting the observed values in the cells of a contingency table in terms of a regression based on sets of observed marginals as explanatory variables.

1. Discrimination information. To make the discussion more specific we shall present it in terms of the analysis of four-way contingency tables. All the essential features of a more general presentation appear. Let us consider the space Ω

of four-way contingency tables $R \times S \times T \times U$ of dimension $r \times s \times t \times u$ so that the generic variable is $\omega = (i,j,k,\ell)$, $i = 1,\ldots,r$, $j = 1,\ldots,s$, $k = 1,\ldots,t$, $\ell = 1,\ldots,u$. Suppose there are two probability distributions or contingency tables (we shall use these terms interchangeably) defined over the space Ω, say $p(\omega)$, $\pi(\omega)$, $\sum_\Omega p(\omega) = 1$, $\sum_\Omega \pi(\omega) = 1$. The discrimination information is defined by

$$(1.1) \quad I(p:\pi) = \sum_\Omega p(\omega) \ln \frac{p(\omega)}{\pi(\omega)} .$$

The basis for this definition, its properties, and relation to other definitions of information measures may be found in [11], in the Proceedings of [13] and references therein. For the particular types of application of interest here the π-distribution, $\pi(\omega)$, in the definition (1.1) according to the problem of interest may either be specified, or it may be an estimated distribution, or it may be an observed distribution. The p-distribution, $p(\omega)$, in the definition (1.1) ranges over or is a member of a family of distributions of interest.

Of the various properties of $I(p:\pi)$ we mention in particular the fact that $I(p:\pi) > 0$ and $= 0$ if and only if $p(\omega) = \pi(\omega)$.

2. Minimum discrimination information estimation. Many problems in the analysis of contingency tables may be characterized as estimating a distribution or contingency table subject to certain restraints and then comparing the estimated table with an

observed table to determine whether the observed table satisfies a null hypothesis implied by the restraints. In accordance with the principle of minimum discrimination information estimation we select that member of the family of p-distributions satisfying the restraints which minimizes the discrimination information $I(p:\pi)$ over the family of pertinent p-distributions. We denote the minimum discrimination information estimate by $p^*(\omega)$ so that

$$(2.1)\quad I(p^*:\pi) = \sum p^*(\omega) \ln \frac{p^*(\omega)}{\pi(\omega)} = \min I(p:\pi).$$

Unless otherwise stated, the summation is over Ω which will be omitted.

In one class of problems the restraints specify some requirement external to the observed values, for example, that a set of marginals have specified values as determined by genetic or other theory [4], [5], [12], or that marginals be homogeneous [3], [14], or that the distribution satisfy certain symmetry conditions [3]. In such problems $\pi(\omega)$ is taken to be an observed contingency table, that is, $x(\omega) = x(ijk\ell) = n\pi(ijk\ell)$, where $n = \sum x(\omega)$.

In another class of problems the restraints specify that the estimated distribution or contingency table have some set of marginals which are the same as those of an observed contingency table. In such cases $\pi(\omega)$ is taken to be either the uniform distribution $\pi(ijk\ell) = 1/rstu$ or a distribution already estimated subject to restraints contained in and implied by the restraints under examination. The latter case includes the classical hypotheses of independence, conditional independence, homogeneity, conditional homogeneity and interaction, all of which can be considered as instances of generalized independence [3], [6], [7], [8], [9], [10], [13], and will be considered in some detail in this paper.

3. <u>Minimum discrimination information statistic</u>. To test whether an observed contingency table satisfies the null hypothesis as represented by the minimum discrimination information estimate we compute a measure of the deviation between the observed distribution and the appropriate estimate by the minimum discrimination information statistic. For notational convenience and later computational convenience let us denote the estimated contingency table in terms of occurrences by $x^*(\omega) = np^*(\omega)$, then for the first category of problems, that is, with restraints determined by external considerations, the minimum discrimination information statistic turns out to be

$$(3.1)\quad 2I(x^*:x) = 2\Sigma\, x^*(\omega)\ \ln \frac{x^*(\omega)}{x(\omega)}$$

which is asymptotically distributed as a χ^2 with appropriate degrees of freedom under the null hypothesis. For the second category of problems, that is, with the restraints implied by a set of observed marginals, or those of a generalized independence hypothesis, the m.d.i. statistic is

$$(3.2)\quad 2I(x:x^*) = 2\Sigma\, x(\omega)\, \ln \frac{x(\omega)}{x^*(\omega)}$$

which is asymptotically distributed as a χ^2 with appropriate degrees of freedom under the null hypothesis.

The statistic in (3.2) is also minus twice the logarithm of the likelihood ratio statistic but this is not true for the statistic in (3.1) or in other applications [11].

4. Minimum discrimination information theorem. We now present a theorem which is the basis for the principle of minimum discrimination information estimation and its applications. We shall present it in a form related to the context of this discussion on the analysis of contingency tables.

Let us consider the space Ω mentioned in section 1 and the discrimination information introduced in (1.1). Suppose now, for example, that we have three linearly independent statistics of interest defined over the space Ω

$$(4.1)\quad T_1(\omega),\ T_2(\omega),\ T_3(\omega).$$

Let us determine the value of $p(\omega)$ which minimizes the discrimination information

$$(4.2)\quad I(p:\pi) = \Sigma\, p(\omega)\, \ln \frac{p(\omega)}{\pi(\omega)}$$

over the family of p-distributions which satisfy the restraints

$$
(4.3)\quad \begin{aligned} \Sigma\, T_1(\omega)\; p(\omega) &= \theta_1^* \\ \Sigma\, T_2(\omega)\; p(\omega) &= \theta_2^* \\ \Sigma\, T_3(\omega)\; p(\omega) &= \theta_3^* \end{aligned}
$$

where θ_1^*, θ_2^*, θ_3^* are specified values.

If $\pi(\omega)$ satisfies the restraints (4.3) then of course the minimum value of $I(p:\pi)$ is zero and the minimizing distribution is $p^*(\omega) = \pi(\omega)$. More generally, the minimum discrimination information theorem [11] states that the minimizing distribution is given by

$$
(4.4)\quad p^*(\omega) = \frac{\exp(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)).\, \pi(\omega)}{M(\tau_1, \tau_2, \tau_3)}
$$

where

$$
(4.5)\quad M(\tau_1, \tau_2, \tau_3) = \Sigma \exp(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega))\; \pi(\omega)
$$

and the τ's are parameters which are in essence undetermined Lagrange multipliers whose values are defined in terms of θ_1^*, θ_2^*, θ_3^* by

$$
\theta_1^* = \frac{\partial}{\partial \tau_1} \ln M(\tau_1, \tau_2, \tau_3) =
$$

$$
= (\Sigma \exp\, (\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)) T_1(\omega) \pi(\omega))/M(\tau_1, \tau_2, \tau_3)
$$

$$(4.6)\quad \theta_2^* = \frac{\partial}{\partial \tau_2} \ln M(\tau_1, \tau_2, \tau_3) =$$

$$= (\Sigma \exp(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)) T_2(\omega)\pi(\omega))/M(\tau_1, \tau_2, \tau_3)$$

$$\theta_3^* = \frac{\partial}{\partial \tau_3} \ln M(\tau_1, \tau_2, \tau_3) =$$

$$= (\Sigma \exp(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)) T_3(\omega)\pi(\omega))/M(\tau_1, \tau_2, \tau_3).$$

We can now state a number of consequences of the preceding.

We note first that $p^*(\omega)$ is a member of an exponential family of distributions generated by $\pi(\omega)$ and as such has the properties of members of an exponential family. In particular $p^*(\omega) = \pi(\omega)$ for $\tau_1 = \tau_2 = \tau_3 = 0$. We may also write (4.4)

$$(4.7)\quad \ln \frac{p^*(\omega)}{\pi(\omega)} = -\ln M(\tau_1, \tau_2, \tau_3) + \tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)$$

$$= L + \tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)$$

with $L = -\ln M(\tau_1, \tau_2, \tau_3)$. The regression expression in (4.7) for $\ln(p^*(\omega)/\pi(\omega))$ with $T_1(\omega)$, $T_2(\omega)$, $T_3(\omega)$ as the explanatory variables and τ_1, τ_2, τ_3 as the regression coefficients plays an important role in the analysis we shall consider.

We note next that the minimum value of the discrimination information (4.2) is

$$(4.8)\quad I(p^*:\pi) = \tau_1 \theta_1^* + \tau_2 \theta_2^* + \tau_3 \theta_3^* - \ln M(\tau_1, \tau_2, \tau_3)$$

where the θ^*'s are defined in (4.3) and the τ's are determined to satisfy (4.6). Using the value in (4.7) it may be shown that if $p(\omega)$ is any member of the family of distributions satisfying (4.3), then

$$(4.9) \quad I(p:\pi) = I(p:p^*) + I(p^*:\pi).$$

The pythagorean property (4.9) plays an important role in the analysis of information tables.

We note thirdly relations connecting the θ^*'s, the τ's, and the covariance matrix of the $T(\omega)$'s. If we define the matrices (vectors)

$$(\underline{d\theta^*})' = (d\theta_1^*,\ d\theta_2^*,\ d\theta_3^*), \quad (\underline{d\tau})' = (d\tau_1,\ d\tau_2,\ d\tau_3)$$

then [11, p.49]

$$(4.10) \quad (\underline{d\theta^*}) = \underline{\Sigma}^*(\underline{d\tau}), \quad (\underline{d\tau}) = \underline{\Sigma}^{*-1}(\underline{d\theta^*})$$

where $\underline{\Sigma}^*$ is the covariance matrix of $T_1(\omega)$, $T_2(\omega)$, $T_3(\omega)$ for the distribution $p^*(\omega)$, that is, with

$$\sigma_{ij}^* = \Sigma(T_i(\omega)-\theta_i^*)(T_j(\omega)-\theta_j^*)p^*(\omega), \ \underline{\Sigma}^* = (\sigma_{ij}^*), \underline{\Sigma}^{*-1} = (\sigma^{*ij})$$

$$(4.11) \quad \frac{\partial\theta_i^*}{\partial\tau_j} = \sigma_{ij}^*, \quad \frac{\partial\tau_i}{\partial\theta_j^*} = \sigma^{*ij}$$

From (4.5) it is seen that $M(\tau_1,\tau_2,\tau_3)$ is the moment-generating function of $T_1(\omega)$, $T_2(\omega)$, $T_3(\omega)$ under the distribution $\pi(\omega)$, hence the cumulant-generating function is given up to quadratic terms by

$$(4.12) \quad \ln M(\tau_1,\tau_2,\tau_3) \approx \theta_1\tau_1 + \theta_2\tau_2 + \theta_3\tau_3 + \tfrac{1}{2}\sum_i\sum_j \sigma_{ij}\,\tau_i\tau_j$$

where

$$(4.13) \quad \theta_i = \Sigma\, T_i(\omega)\,\pi(\omega),\ \sigma_{ij} = \Sigma(T_i(\omega) - \theta_i)(T_j(\omega) - \theta_j)\,\pi(\omega).$$

Thus, using (4.12) in (4.6), we get

$$(4.14) \quad \begin{aligned} \theta_1^* &\approx \theta_1 + \sum_j \sigma_{1j}\,\tau_j \\ \theta_2^* &\approx \theta_2 + \sum_j \sigma_{2j}\,\tau_j \\ \theta_3^* &\approx \theta_3 + \sum_j \sigma_{3j}\,\tau_j \end{aligned}$$

and then using (4.14) in (4.8) yields

$$(4.15) \quad 2I(p^*:\pi) \approx (\underline{\theta}^* - \underline{\theta})'\,\underline{\Sigma}^{-1}\,(\underline{\theta}^* - \underline{\theta}) \approx \underline{\tau}'\,\underline{\Sigma}\,\underline{\tau}.$$

We have used three functions $T_1(\omega)$, $T_2(\omega)$, $T_3(\omega)$ thus far in the discussion merely as a matter of convenience. We note that (4.15) holds for a set of m functions $T_i(\omega)$, $i = 1,\ldots,m$ with appropriate meanings for the matrices. Let us partition the set of m functions $T_i(\omega)$ into a set H_a say of m_a and a set H_c of the remaining $m_c = m - m_a$ functions, where the functions in the set H_a have the property that

$$(4.16) \quad \theta_i^* = \theta_i \,, \quad i = 1,\ldots,m_a \;.$$

We have the related partitioning of the covariance matrix of the $T_i(\omega)$, $i = 1,\ldots,m$

$$(4.17) \quad \underline{\Sigma} = \begin{pmatrix} \underline{\Sigma}_{aa} & \underline{\Sigma}_{ab} \\ \underline{\Sigma}_{ba} & \underline{\Sigma}_{bb} \end{pmatrix}, \quad \underline{\Sigma}_{ba} = \underline{\Sigma}'_{ab}$$

and the $\underline{\theta}$, $\underline{\theta}^*$, and $\underline{\tau}$ matrices

$$(4.18) \quad \underline{\theta}^{*\prime} = (\underline{\theta}_a^{*\prime}, \underline{\theta}_b^{*\prime}), \quad \underline{\theta}' = (\underline{\theta}_a', \underline{\theta}_b'), \quad \underline{\tau}' = (\underline{\tau}_a', \underline{\tau}_b').$$

In terms of the partitionings in (4.17) and (4.18) the relations in (4.14) may be written as

$$(4.19) \quad \begin{aligned} \underline{\theta}_a^* &\approx \underline{\theta}_a + \underline{\Sigma}_{aa}\,\underline{\tau}_a + \underline{\Sigma}_{ab}\,\underline{\tau}_b \\ \underline{\theta}_b^* &\approx \underline{\theta}_b + \underline{\Sigma}_{ba}\,\underline{\tau}_a + \underline{\Sigma}_{bb}\,\underline{\tau}_b \end{aligned}$$

and using the fact that $\underline{\theta}_a^* = \underline{\theta}_a$, it is found that using these results in (4.8) now yields

$$(4.20) \quad 2I(p^*:\pi) \approx (\underline{\theta}_b^* - \underline{\theta}_b)'\, \underline{\Sigma}_{bb.a}^{-1}\, (\underline{\theta}_b^* - \underline{\theta}_b) \approx \underline{\tau}_b'\, \underline{\Sigma}_{bb.a}\,\underline{\tau}_b$$

where $\underline{\Sigma}_{bb.a} = \underline{\Sigma}_{bb} - \underline{\Sigma}_{ba}\,\underline{\Sigma}_{aa}^{-1}\,\underline{\Sigma}_{ab}$ is an $m_b \times m_b$ matrix. The results under the partitioning will help in interpreting the analysis of information values and are similar to those occurring in the testing of subhypotheses in the linear and multivariate linear hypothesis theory [11, p. 216, 259].

We note from (4.6) and (4.7) that

$$(4.21)\quad \frac{\partial}{\partial \tau_i} \ln p^*(\omega) = T_i(\omega) - \frac{\partial}{\partial \tau_i} \ln M(\tau_1, \tau_2, \ldots) = T_i(\omega) - \theta_i^*,$$

hence $T_i(\omega)$ is the maximum likelihood estimator of θ_i^*. Thus if we write $T_i(\omega) = \hat{\theta}_i^*$ and denote the values satisfying (4.6) or (4.14) with $\hat{\theta}_i^*$ in place of θ_i^* and $\hat{\tau}_i$ in place of τ_i, we have corresponding to (4.15)

$$(4.22)\quad \begin{aligned} 2I(\hat{p}^* : \pi) &= 2 \sum_i \hat{\tau}_i \hat{\theta}_i^* - 2 \ln M(\hat{\tau}_1, \hat{\tau}_2, \ldots) \\ &\approx (\underline{\hat{\theta}}^* - \underline{\theta})' \underline{\Sigma}^{-1} (\underline{\hat{\theta}}^* - \underline{\theta}) \approx \underline{\hat{\tau}}' \underline{\Sigma}\, \underline{\hat{\tau}} \end{aligned}$$

and corresponding to (4.20)

$$(4.23)\quad 2I(\hat{p}^* : \pi) = (\underline{\hat{\theta}}_o^* - \underline{\theta}_o)' \underline{\Sigma}_{o \cdot a}^{-1} (\underline{\hat{\theta}}_o^* - \underline{\theta}_o) \approx \underline{\hat{\tau}}_o' \underline{\Sigma}_{o \cdot a} \underline{\hat{\tau}}_o .$$

We remark that the covariance matrix of the $\hat{\tau}$'s is the inverse of the covariance matrix of the $T_i(\omega)$'s.

If the $\hat{\theta}_i^*$ are the averages of n independent observations then we have for the minimum discrimination information statistics

$$(4.24)\quad 2n\, I(\hat{p}^* : \pi) \approx n(\underline{\hat{\theta}}^* - \underline{\theta})' \underline{\Sigma}^{-1} (\underline{\hat{\theta}}^* - \underline{\theta}) \approx n\, \underline{\hat{\tau}}' \underline{\Sigma}\, \underline{\hat{\tau}}$$

and in the partitioned case

$$(4.25)\quad 2n\, I(\hat{p}^* : \pi) \approx n(\underline{\hat{\theta}}_o^* - \underline{\theta}_o)' \underline{\Sigma}_{o \cdot a}^{-1} (\underline{\hat{\theta}}_o^* - \underline{\theta}_o) \approx n\, \underline{\hat{\tau}}_o' \underline{\Sigma}_{o \cdot a} \underline{\hat{\tau}}_o .$$

Under the null hypothesis $2n\, I(\hat{p}^* : \pi)$ in (4.24) or (4.25) is asymptotically distributed as χ^2 respectively with m or m_o degrees of freedom.

5. Computational procedures. An experiment has been designed and observations made resulting in a multidimensional contingency table with the desired classifications and categories. All the information the experimenter hopes to obtain from the experiment is contained in the contingency table. In the process of analysis, the aim is to express the observed table by a number of parameters depending on some or all of the marginals, that is, to find out how much of this total information is contained in a summary consisting of sets of marginals. Indeed, the relationship between the concept of independence or association and interaction in contingency tables and the role the marginals play is evidenced in the writings of Bartlett [1], Simpson [17], Roy and Kastenbaum [16], Lewis [15], Darroch [2] and others on the analysis of contingency tables. Thus, the θ's in the preceding discussion will be the marginals of interest.

5.1. The $T(\omega)$ functions. The $T(\omega)$ functions for the $R \times S \times T \times U$ table turn out to be a basic set of simple functions and their various products. Thus, for example, the $T(\omega)$ function associated with the one-way marginal $p(2\ldots)$ is

$$(5.1)\quad T_2^R(ijk\ell) = 1 \quad \text{for } i = 2, \text{ any } j,k,\ell$$
$$= 0 \quad \text{otherwise}$$

since

$$(5.2)\quad \Sigma\, p(ijk\ell)\, T_s^R(ijk\ell) = p(2\ldots).$$

Similarly the $T(\omega)$ function associated with the one-way marginal $p(..3.)$, for example, is

$$(5.3)\quad \begin{aligned} T_s^T(ijk\ell) &= 1 \qquad \text{for } k = 3, \text{ any } i,j,\ell \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

since

$$(5.4)\quad \Sigma\, p(ijk\ell)\, T_s^T(ijk\ell) = p(..3.).$$

Thus for the $r \times s \times t \times u$ table we have

$$(5.5)\quad \begin{aligned} &(r-1) \text{ linearly independent functions } T_\alpha^R(ijk\ell), \alpha=1,\ldots,r-1 \\ &(s-1) \text{ linearly independent functions } T_\beta^S(ijk\ell), \beta=1,\ldots,s-1 \\ &(t-1) \text{ linearly independent functions } T_\gamma^T(ijk\ell), \gamma=1,\ldots,t-1 \\ &(u-1) \text{ linearly independent functions } T_\delta^U(ijk\ell), \delta=1,\ldots,u-1, \end{aligned}$$

since, for example,

$$\sum_{\alpha=1}^{r} \Sigma\, T_\alpha^R(ijk\ell) = rstu\ .$$

We have arbitrarily excluded the functions corresponding to $\alpha = r$, $\beta = s$, $\gamma = t$, $\delta = u$ as a matter of convenience, we could have selected $\alpha = 1$, $\beta = 1$, $\gamma = 1$, $\delta = 1$ or any other set of values.

The $T(\omega)$ function associated with the two-way marginal $p(12..)$ say, is $T_1^R(ijk\ell)\ T_2^S(ijk\ell)$ since from the definition of $T_1^R(ijk\ell)$ and $T_2^S(ijk\ell)$ it may be seen that

$$(5.6)\quad T_1^R(ijk\ell)\ T_2^S(ijk\ell) = 1 \quad \text{for } i = 1,\ j = 2, \text{ any } k,\ell$$
$$= 0 \quad \text{otherwise}$$

and

$$(5.7)\quad \Sigma\ p(ijk\ell)\ T_1^R(ijk\ell)\ T_2^S(ijk\ell) = p(12..).$$

Thus the $T(\omega)$ function associated with any two-way marginal is a product of two appropriate functions of the set (5.5).

Similarly the $T(\omega)$ function associated with any three-way marginal will be a product of three of the appropriate functions of the set (5.5), for example,

$$(5.8)\quad \Sigma\ p(ijk\ell)\ T_2^R(ijk\ell)\ T_1^T(ijk\ell)\ T_3^U(ijk\ell) = p(2.13).$$

Similarly the $T(\omega)$ function associated with any four-way marginal will be a product of four of the appropriate functions of the set (5.5), for example,

$$(5.9)\quad \Sigma\ p(ijk\ell)\ T_2^R(ijk\ell)\ T_1^S(ijk\ell)\ T_1^T(ijk\ell)\ T_2^U(ijk\ell) = p(2112).$$

We note that there are a total of

$N_1 = (r-1) + (s-1) + (t-1) + (u-1)$

$N_2 = (r-1)(s-1)+(r-1)(t-1)+(r-1)(u-1)+(s-1)(t-1)+(s-1)(u-1)+(t-1)(u-1)$

$N_3 = (r-1)(s-1)(t-1)+(r-1)(s-1)(u-1)+(r-1)(t-1)(u-1)+(s-1)(t-1)(u-1)$

$N_4 = (r-1)(s-1)(t-1)(u-1)$

respectively of the simple linearly independent functions and their products two, three, four at a time. It may be verified that

$$(5.10) \quad rstu - 1 = N = N_1 + N_2 + N_3 + N_4 .$$

These values are degrees of freedom in the analysis of information tables in [6], [10].

5.2. The $p^*(\omega)$ values. In the usual regression analysis procedure, one first computes the regression coefficients and then gets the values of the estimates. In this case however we reverse the procedure. Instead of trying to obtain the values of the τ's from (4.6) we shall first obtain the values of $p^*(\omega)$ by a straightforward convergent iterative procedure and then derive the values of the τ's from (4.7). We shall not discuss the details of the iteration here since they have been described in [4], [6], [9], [10]. The iteration may be described as successively cycling through adjustments of the marginals of interest starting with the $\pi(\omega)$ distribution until a desired accuracy of agreement between the set of observed marginals of interest and the computed marginals has been attained.

5.3. The τ values. From the definitions of the $T(\omega)$ functions in section 5.1 it is clear that they take on only the values 0 or 1 for each value of ω. From the nature of the $T(\omega)$ functions the set of regression equations (4.7) will have some with a single τ value which can be determined. Then there will be a set with one additional unknown value and some of the τ's already determined. These new unknown τ values can be then determined. This process of successive evaluation is carried on until all the values of τ are determined.

6. Analysis of information. Although the preceding theoretical discussion has been in terms of probabilities, estimated probabilities or relative frequencies, in practice it has been found more convenient not to divide everything by n, the total number of occurrences, and deal with observed or estimated occurrences, that is, with $n\pi(ijk\ell) = n/rstu$, $x(ijk\ell)$, $x(i...)$, $x(.jk.)$, $x^*(ijk\ell) = n\,p^*(ijk\ell)$ etc. The analysis of information is based on the fundamental relation (4.9) for the minimum discrimination information statistics. Specifically if $n\,p_a^*(\omega) = x_a^*(\omega)$ is the minimum discrimination information estimate corresponding to a set H_a of given marginals and $x_b^*(\omega)$ is the minimum discrimination information estimate corresponding to a set H_b of given marginals, where $H_a \subset H_b$, then the basic relations are

$$
(6.1)\quad
\begin{aligned}
2I(x:n\pi) &= 2I(x_a^*:n\pi) + 2I(x:x_a^*)\\
2I(x:n\pi) &= 2I(x_b^*:n\pi) + 2I(x:x_b^*)\\
2I(x_b^*:n\pi) &= 2I(x_a^*:n\pi) + 2I(x_b^*:x_a^*)\\
2I(x:x_a^*) &= 2I(x_b^*:x_a^*) + 2I(x:x_b^*)
\end{aligned}
$$

In terms of the representation in (4.4) as an exponential family, for our discussion, the two extreme cases are the uniform distribution for which all τ's are zero, and the observed contingency table or distribution for which all N = rstu - 1 τ's are needed.

Measures of the form $2I(x:x_a^*)$, that is, the comparison of an observed contingency table with an estimated contingency table, are called measures of interaction and measures of the form $2I(x_b^*:x_a^*)$, that is, the comparison of two estimated contingency tables, are called measures of effect, that is the effect of the marginals in the set H_b but not in the set H_a. From the results in (4.24) or (4.25) we see that $2I(x:x_a^*)$ tests a null hypothesis that the set of τ parameters in the representation of the observed contingency table $x(\omega)$ but not in the representation of the estimated table $x_a^*(\omega)$ are zero, and $2I(x_b^*:x_a^*)$ tests a null hypothesis that the additional set of τ parameters in the representation of the estimated table $x_b^*(\omega)$ but not in the representation of the estimated table $x_a^*(\omega)$ are zero.

Since the marginals of the estimated table $x_a^*(\omega)$ which form the set of restraints H_a used to generate $x_a^*(\omega)$ are the same as the corresponding marginals of the observed $x(\omega)$ table and all lower order implied marginals, $2I(x:x_a^*)$ is also approximately a quadratic in the differences between the remaining marginals of the $x(\omega)$ table and the corresponding ones as calculated from the $x_a^*(\omega)$ table.

Similarly $2I(x_b^*:x_a^*)$ is also approximately a quadratic in the differences between those additional marginals in H_b but not in H_a and the corresponding marginal values as computed from the

$x^*_{a}(\omega)$ table.

As we shall see, because of the nature of the $T(\omega)$ functions described in section 5.1, the τ's are determined from the regression equations (4.7) as sums and differences of values of $\ln x^*(ijk\ell)$. A variety of statistics have been presented in the literature for the analysis of contingency tables which are quadratics in the marginal values or quadratics in the logarithms of the observed or estimated values. The principle of minimum discrimination information estimation and its procedures thus provides a unifying relationship since such statistics may be seen as opposite faces of the minimum discrimination information statistic.

We have presented the approximations in terms of quadratic forms in the marginals or the τ's to assist in understanding and interpreting the analysis of information tables as a bridge connecting the familiar procedures of classical regression analysis and the procedures proposed here. The covariance matrix of the $T(\omega)$ functions can be estimated for either the observed table or any of the estimated tables and the inverse of that matrix found should their values be desired.

7. The 2×2 table. Before we present an application of the preceding ideas to experimental data in a four-way contingency table, we shall reexamine the 2×2 table from the point of view of this paper. The algebraic details are simple in this case and exhibit the unification of the information theoretic development.

Suppose we have the observed 2×2 table in figure 7.1.

If we fit the one-way

x(11)	x(12)	x(1.)
x(21)	x(22)	x(2.)
x(.1)	x(.2)	n

Figure 7.1

marginals, the generalized independence hypothesis is the classical independence hypothesis and the minimum discrimination information estimate is $x^*(ij) = x(i.)x(.j)/n$. A convenient representation of the regression (4.7) is given in figure 7.2. The entries in the columns τ_1, τ_2, τ_3

i	j	L	τ_1	τ_2	τ_3
1	1	1	1	1	1
1	2	1	1		
2	1	1		1	
2	2	1			

Figure 7.2

are respectively the values of the functions $T_1(ij)$, $T_2(ij)$, $T_3(ij)$ associated with the marginals $\theta_1 = x(1.)$, $\theta_2 = x(.1)$, $\theta_3 = x(11)$, and the column headed L corresponds to the negative of the logarithm of the moment-generating function. For the observed distribution, recalling the regression (4.7), it is found that

$$(7.1)\quad L = \ln\,(x(22)/n/4),\ \tau_1 = \ln\,(x(12)/x(22)),\ \tau_2 = \ln\,(x(21)/x(22))$$
$$\tau_3 = \ln\,(x(11)x(22)/x(12)x(21)).$$

If we call $\underline{T}$ the matrix with columns the columns of Figure 7.2, that is,

$$(7.2)\quad \underline{T} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

and define a diagonal matrix $\underline{D}$ with main diagonal the elements $x(ij)$, that is,

$$(7.3)\quad \underline{D} = \begin{pmatrix} x(11) & 0 & 0 & 0 \\ 0 & x(12) & 0 & 0 \\ 0 & 0 & x(21) & 0 \\ 0 & 0 & 0 & x(22) \end{pmatrix}$$

then it may be verified that the estimate of the covariance matrix of the $T_i(\omega)$ for the observed contingency table is $\underline{\Sigma} = \underline{A}_{22.1}$ where

$$(7.4)\quad \underline{A} = \begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix} = \underline{T}'\underline{DT}$$

$$(7.5)\quad \underline{A}_{22.1} = \underline{A}_{22} - \underline{A}_{21}\,\underline{A}_{11}^{-1}\,\underline{A}_{12}$$

and $\underline{A}_{11}$ is 1×1, $\underline{A}_{22}$ is 3×3, $\underline{A}'_{21} = \underline{A}_{12}$ is 1×3. It is found that

$$(7.6)\quad \underline{\Sigma} = \begin{pmatrix} \dfrac{x(1.)x(2.)}{n} & x(11) - \dfrac{x(1.)x(.1)}{n} & \dfrac{x(11)x(2.)}{n} \\ x(11) - \dfrac{x(1.)x(.1)}{n} & \dfrac{x(.1)x(.2)}{n} & \dfrac{x(11)x(.2)}{n} \\ \dfrac{x(11)x(2.)}{n} & \dfrac{x(11)x(.2)}{n} & x(11) - \dfrac{x^2(11)}{n} \end{pmatrix}$$

Even for this simple case inverting the matrix in (7.6) is messy algebraically, however, it is easier to use the relations in (4.10) and (4.11). We have from (7.1)

$$(7.7)\qquad \tau_1 = \ell n\ x(12) - \ell n\ x(22),\quad \tau_2 = \ell n\ x(21) - \ell n\ x(22),$$
$$\tau_3 = \ell n\ x(11) + \ell n\ x(22) - \ell n\ x(12) - \ell n\ x(21)$$

and from $\theta_1 = x(1.)$, $\theta_2 = x(.1)$, $\theta_3 = x(11)$ and the relations implied in Figure 7.1 it is found that

$$(7.8)\qquad x(11) = \theta_3,\ x(12) = \theta_1 - \theta_3,\ x(21) = \theta_2 - \theta_3,\ x(22) = n - \theta_1 - \theta_2 + \theta_3 .$$

It then follows that

$$(7.9)\qquad \begin{aligned}
&\frac{\partial\tau_1}{\partial\theta_1} = \frac{1}{x(12)} + \frac{1}{x(22)},\quad \frac{\partial\tau_1}{\partial\theta_2} = \frac{1}{x(22)},\quad \frac{\partial\tau_1}{\partial\theta_3} = -\frac{1}{x(12)} - \frac{1}{x(22)}\\
&\frac{\partial\tau_2}{\partial\theta_1} = \frac{1}{x(22)},\quad \frac{\partial\tau_2}{\partial\theta_2} = \frac{1}{x(21)} + \frac{1}{x(22)},\quad \frac{\partial\tau_2}{\partial\theta_3} = -\frac{1}{x(21)} - \frac{1}{x(22)}\\
&\frac{\partial\tau_3}{\partial\theta_1} = -\frac{1}{x(12)} - \frac{1}{x(22)},\quad \frac{\partial\tau_3}{\partial\theta_2} = -\frac{1}{x(21)} - \frac{1}{x(22)},\\
&\qquad\frac{\partial\tau_3}{\partial\theta_3} = \frac{1}{x(11)} + \frac{1}{x(12)} + \frac{1}{x(21)} + \frac{1}{x(22)}
\end{aligned}$$

that is, the entries of $\underline{\Sigma}^{-1}$ since $\dfrac{\partial\tau_i}{\partial\theta_j} = \sigma^{ij}$.

Note that the value of the logarithm of the cross-product ratio as a measure of association appears in the course of the analysis as the value of τ_3, and that $\tau_3 = 0$ for $x^*(ij)$ whose representation as in Figure 7.2 does not involve the last column. The minimum discrimination information statistic to test the null hypothesis of independence is $2I(x:x^*)$. In this case $\theta_1^* = \theta_1$, $\theta_2^* = \theta_2$ and in accordance with (4.25)

$$(7.10)\quad 2I(x:x^*) \approx \left(x(11) - \frac{x(1.)x(.1)}{n}\right)^2 \left(\frac{1}{x^*(11)}+\frac{1}{x^*(12)}+\frac{1}{x^*(21)}+\frac{1}{x^*(22)}\right).$$

Remembering that $x^*(ij) = x(i.)x(.j)/n$, the right-hand side of (7.10) may also be shown to be

$$(7.11)\quad \sum \left(x(ij) - x(i.)x(.j)/n\right)^2 \Big/ \frac{x(i.)x(.j)}{n}$$

the classical X^2-test for independence with one degree of freedom. A test which has been proposed for the null hypothesis of no association or no interaction in the 2×2 table is

$$(7.12)\quad (\ln x(11) + \ln x(22) - \ln x(12) - \ln x(21))^2 \left(\frac{1}{x(11)}+\frac{1}{x(12)}+\frac{1}{x(21)}+\frac{1}{x(22)}\right)^{-1}$$

which is seen to be the approximation for $2I(x:x^*)$ in terms of the τ's with the covariance matrix estimated using the observed values and not the estimated values. We remark that if the observed values are used to estimate the covariance matrix then instead of the classical X^2- test in (7.11) there is derived the modified

Neyman chi-square

$$(7.13) \quad X'^2 = \Sigma(x(ij) - x(i.)x(.j)/n)^2/x(ij).$$

8. Example with experimental data. Consider the $R \times S \times T \times U$ table 8.1a representing the results of test shooting under three different conditions:

R:Gun barrel wear:i=1, new, i=2, moderate, i=3, excessive

S:Gun barrel temperature:j=1, cold, j=2, hot

T:Unit temperature:k=1, hot, k=2, ambient, k=3, cold

U:Number operative:ℓ=1, success, ℓ=2, failure.

We are indebted to Mr. B.M. Kurkjian of the Harry Diamond Laboratories for the data and his interest in the analytic procedure we have discussed. We note that 15 rounds each were fired under each of 18 experimental conditions. This is not necessary for the application of the analysis of information procedures but was required for the earlier application of Brandt's analysis to the data.

Figure 8.1 presents a graphic representation of the regression (4.7) and is similar to that in Figure 7.2 for the 2 x 2 table. The L column corresponds to the negative of the logarithm of the moment-generating function (a normalizing value) and each of the other columns is a $T(\omega)$ function with the associated τ value at the head of the column. Superscripts and subscripts are used to identify the factors and categories involved. The complete representation in Figure 8.1 with the 35 τ values will provide an exact representation for the observed values $x(\omega)$. Tables 8.2, 8.3, and 8.4 are analysis of information tables presenting appropriate analyses as various sets of marginals of interest are introduced as explanatory variables.

In Figure 8.2 the columns corresponding to the τ parameters which enter into the various distributions appearing in tables 8.2, 8.3 and 8.4 have been checked. Note that for $n\pi$, the uniform distribution, there are no checks, and for $x(\omega)$, the observed distribution all columns are checked. The degrees of freedom for any effect component is the difference in the number of columns checked for the corresponding estimates. The degrees of freedom for any interaction component is the difference in the number of columns checked for the observed x-distribution and the estimated distribution.

The null hypothesis for any effect component is that the additional τ parameters are zero, for example, the null hypothesis for the effect component $2I(x^*_a:x^*_b)$ in table 8.2 is that τ^{RU}_{11}, τ^{RU}_{21} are zero. The null hypothesis for any interaction component is that the set of parameters which are checked for the observed x-distribution but not for the estimated distribution are zero, for example, the null hypothesis for the third-order interaction component $2I(x:x^*)$ in table 8.2 is that τ^{RSTU}_{1111}, τ^{RSTU}_{2111}, τ^{RSTU}_{1121}, τ^{RSTU}_{2121} are zero.

Note that all the marginals implied for x^* in table 8.4 are $x(i\ldots)$, $x(.j..)$, $x(..k.)$, $x(\ldots\ell)$, $x(ij..)$, $x(i.k.)$, $x(i..\ell)$, $x(.jk.)$, $x(.j.\ell)$, $x(ijk.)$, $x(ij.\ell)$ and the marginals implied for $x^{*\prime}$ in table 8.4 are $x(i\ldots)$, $x(.j..)$, $x(..k.)$, $x(\ldots\ell)$, $x(ij..)$, $x(i.k.)$, $x(i..\ell)$, $x(.jk.)$, $x(.j.\ell)$, $x(..k\ell)$, $x(ijk.)$, $x(ij.\ell)$, $x(i.k\ell)$, hence the six parameters $\tau^{TU}_{11}, \tau^{TU}_{21}, \tau^{RTU}_{111}, \tau^{RTU}_{211}, \tau^{RTU}_{121}, \tau^{RTU}_{221}$ appear in x^* but not in x^*.

We draw the following conclusions from tables 8.2, 8.3, 8.4:

1. Success/Failure is not homogeneous over the 18 experimental situations, $x_a^*(ijk\ell) = x(ijk.)x(...\ell)/n$, $2I(x:x_a^*) = 34.371$, 17 D.F.

2. The effect of $x(i..\ell)$ in table 8.2 is almost significant, but those of $x(.j.\ell)$, $x(..k\ell)$ are not significant, hence we proceed as in table 8.4.

3. The marginals $x(ijk.)$, $x(ij.\ell)$, $x(i.k\ell)$ and the lower order marginals they imply provide an acceptable estimate for the original data since $2I(x:x_a^*) = 7.413$, 6 D.F., that is, we accept a null hypothesis that the set of six parameters τ_{111}^{STU}, τ_{121}^{STU}, τ_{1111}^{RSTU}, τ_{2111}^{RSTU}, τ_{1121}^{RSTU}, τ_{2121}^{RSTU} are zero.

4. Using Figure 8.1 and Figure 8.2 we can express the logarithm of the ratio of the estimates for success to failure under all 18 experimental conditions, that is, the logit, as the linear combination of a constant term τ_1^U, a term depending on barrel wear τ_{11}^{RU}, τ_{21}^{RU}, a term depending on the interaction of barrel wear and barrel temperature τ_{111}^{RSU}, τ_{211}^{RSU}, and a term depending on the interaction of barrel wear and unit temperature τ_{111}^{RTU}, τ_{211}^{RTU}, τ_{121}^{RTU}, τ_{221}^{RTU}.

$$\ln \frac{x_a^*(1111)}{x_a^*(1112)} = \tau_1^U + \tau_{11}^{RU} + \tau_{11}^{SU} + \tau_{11}^{TU} + \tau_{111}^{RSU} + \tau_{111}^{RTU}$$

$$\ln \frac{x_a^*(1211)}{x_a^*(1212)} = \tau_1^U + \tau_{11}^{RU} + \tau_{11}^{TU} + \tau_{111}^{RTU}$$

$$\ln \frac{x_a^*(2111)}{x_a^*(2112)} = \tau_1^U + \tau_{21}^{RU} + \tau_{11}^{SU} + \tau_{11}^{TU} + \tau_{211}^{RSU} + \tau_{211}^{RTU}$$

$$\ln \frac{x_a^*(2211)}{x_a^*(2212)} = \tau_1^U + \tau_{21}^{RU} + \tau_{11}^{TU} + \tau_{211}^{RTU}$$

$$\ln \frac{x_a^*(3111)}{x_a^*(3112)} = \tau_1^{U} + \tau_{11}^{SU} + \tau_{11}^{TU}$$

$$\ln \frac{x_a^*(3211)}{x_a^*(3212)} = \tau_1^{U} + \tau_{11}^{TU}$$

$$\ln \frac{x_a^*(1121)}{x_a^*(1122)} = \tau_1^{U} + \tau_{11}^{RU} + \tau_{11}^{SU} + \tau_{21}^{TU} + \tau_{111}^{RSU} + \tau_{121}^{RTU}$$

$$\ln \frac{x_a^*(1221)}{x_a^*(1222)} = \tau_1^{U} + \tau_{11}^{RU} + \tau_{21}^{TU} + \tau_{121}^{RTU}$$

$$\ln \frac{x_a^*(2121)}{x_a^*(2122)} = \tau_1^{U} + \tau_{21}^{RU} + \tau_{11}^{SU} + \tau_{21}^{TU} + \tau_{211}^{RSU} + \tau_{221}^{RTU}$$

$$\ln \frac{x_a^*(2221)}{x_a^*(2222)} = \tau_1^{U} + \tau_{21}^{RU} + \tau_{21}^{TU} + \tau_{221}^{RTU}$$

$$\ln \frac{x_a^*(3121)}{x_a^*(3122)} = \tau_1^{U} + \tau_{11}^{SU} + \tau_{21}^{TU}$$

$$\ln \frac{x_a^*(3221)}{x_a^*(3222)} = \tau_1^{U} + \tau_{21}^{TU}$$

$$\ln \frac{x_a^*(1131)}{x_a^*(1132)} = \tau_1^{U} + \tau_{11}^{RU} + \tau_{11}^{SU} + \tau_{111}^{RSU}$$

$$\ln \frac{x_a^*(1231)}{x_a^*(1232)} = \tau_1^{U} + \tau_{11}^{RU}$$

$$\ln \frac{x_a^*(2131)}{x_a^*(2132)} = \tau_1^{U} + \tau_{21}^{RU} + \tau_{11}^{SU} + \tau_{211}^{RSU}$$

$$\ln \frac{x_a^*(2231)}{x_a^*(2232)} = \tau_1^{U} + \tau_{21}^{RU}$$

$$\ell n \frac{x_s^*(3131)}{x_s^*(3132)} = \tau_1^U + \tau_{11}^{SU}$$

$$\ell n \frac{x_s^*(3231)}{x_s^*(3232)} = \tau_1^U$$

5. Since the computer program provides not only the values of $x_s^*(ijk\ell)$ but also the values of $\ell n\ (x_s^*(ijk\ell)/x_s^*(3232))$, the values of the τ's in conclusion 4 above can be easily found.

$$\tau_1^U = \ell n \frac{x_s^*(3231)}{x_s^*(3232)} = 3.0281$$

$$\tau_{11}^{RU} = \ell n \frac{x_s^*(1231)}{x_s^*(1232)} - \tau_1^U = -1.6470$$

$$\tau_{21}^{RU} = \ell n \frac{x_s^*(2231)}{x_s^*(2232)} - \tau_1^U = -2.2870$$

$$\tau_{11}^{SU} = \ell n \frac{x_s^*(3131)}{x_s^*(3132)} - \tau_1^U = -0.6794$$

$$\tau_{11}^{TU} = \ell n \frac{x_s^*(3211)}{x_s^*(3212)} - \tau_1^U = -1.9759$$

$$\tau_{21}^{TU} = \ell n \frac{x_s^*(3221)}{x_s^*(3222)} - \tau_1^U = -0.7746$$

$$\tau_{111}^{RSU} = \ell n \frac{x_s^*(1131)}{x_s^*(1132)} - \tau_1^U - \tau_{11}^{RU} - \tau_{11}^{SU} = -0.2928$$

$$\tau_{211}^{RSU} = \ell n \frac{x_s^*(2131)}{x_s^*(2132)} - \tau_1^U - \tau_{21}^{RU} - \tau_{11}^{SU} = 1.7215$$

$$\tau_{111}^{RTU} = \ell n \frac{x_s^*(1211)}{x_s^*(1212)} - \tau_1^U - \tau_{11}^{RU} - \tau_{11}^{TU} = 2.3336$$

$$\tau_{211}^{RTU} = \ell n \frac{x_s^*(2211)}{x_s^*(2212)} - \tau_1^U - \tau_{21}^{RU} - \tau_{11}^{TU} = 1.4528$$

$$\tau_{121}^{RTU} = \ln \frac{x_a^*(1221)}{x_a^*(1222)} - \tau_1^U - \tau_{11}^{RU} - \tau_{21}^{TU} = 0.1639$$

$$\tau_{221}^{RTU} = \ln \frac{x_a^*(2221)}{x_a^*(2222)} - \tau_1^U - \tau_{21}^{RU} - \tau_{21}^{TU} = 0.5878$$

As a check we have, for example, $\ln(x_a^*(1111)/x_a^*(1112)) = 0.7666$ and $\tau_1^U + \tau_{11}^{RU} + \tau_{11}^{SU} + \tau_{11}^{TU} + \tau_{111}^{RSU} + \tau_{111}^{RTU} = 0.7666$.

6. The values of L and other τ parameters for the x_a^*-distribution can be obtained from Figure 8.1 and the computer listing of the values of $\ln(x_a^*(ijk\ell)/x_a^*(3232))$ and $\ln(x_a^*(3232)/n\pi)$, in this case $n\pi = 270/(3\times2\times3\times2)$. Thus L = -2.3822, $\tau_1^R = \ln(x_a^*(1232)/x_a^*(3232)) = 1.4701$, etc.

7. The computer output for $x_a^*(ijk\ell)$ is listed as table 8.5. Five values are given for each i,j,k,ℓ, these are:

Observed: $x(ijk\ell)$

Predicted: $x_a^*(ijk\ell)$

Residual: $x(ijk\ell) - x_a^*(ijk\ell)$

Standardize: $2\,x(ijk\ell)\ \ln\,(x(ijk\ell)/x_a^*(ijk\ell))$

Log ratio: $\ln(x_a^*(ijk\ell)/x_a^*(3232))$.

There is also given the value of $2I(x:x_a^*)$ along with the degrees of freedom and a probability based on the χ^2- distribution and the value of L as log(x STAR/N/CELLS).

9. Acknowledgment. The interest and cooperation of Professor C.T. Ireland and Dr. H.H. Ku are gratefully acknowledged.

Table 8.1a Original Data $x(ijk\ell)$

	i	1						2						3						
	j	1			2			1			2			1			2			
	k	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	
ℓ	1	9	8	9	14	9	12	11	14	12	9	8	11	9	13	13	11	13	15	200
	2	6	7	6	1	6	3	4	1	3	6	7	4	6	2	2	4	2	0	70
		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	270

Table 8.1b Estimated Data $x^{*}_{2}(ijk\ell)$

Marginals Fitted: $x(ijk.)$, $x(ij.\ell)$, $x(i.k\ell)$

	i	1						2					
	j	1			2			1			2		
	k	1	2	3	1	2	3	1	2	3	1	2	3
ℓ	1	10.242	6.746	9.012	12.758	10.254	11.988	11.686	12.473	12.841	8.314	9.527	10.159
	2	4.758	8.254	5.988	2.242	4.746	3.012	3.314	2.527	2.159	6.686	5.473	4.841
		15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000

	i	3						
	j	1			2			
	k	1	2	3	1	2	3	
ℓ	1	8.883	12.425	13.692	11.117	13.575	14.308	200.000
	2	6.118	2.574	1.307	3.882	1.426	0.693	70.000
		15.001	14.999	14.999	14.999	15.001	15.001	270.000

Analysis of Information

Table 8.2

Component due to	Information	D.F.
a) $x(ijk.)$	$2I(x:x_a^*)=99.639$	18
b) $x(ijk.),x(...\ell)$		
U-effect	$2I(x_b^*:x_a^*)=65.268$	1
Interaction	$2I(x:x_b^*)=34.371$	17
c) $x(ijk.),x(i..\ell)$		
RU-effect | RST	$2I(x_c^*:x_b^*)=5.303$	2
Interaction	$2I(x:x_c^*)=29.068$	15
d) $x(ijk.),x(i..\ell),x(.j.\ell)$		
SU-effect | RST, RU	$2I(x_d^*:x_c^*)=0.314$	1
Interaction	$2I(x:x_d^*)=28.754$	14
e) $x(ijk.),x(i..\ell),x(.j.\ell),x(..k\ell)$		
TU-effect | RST, RU, SU	$2I(x_e^*:x_d^*)=2.705$	2
Interaction	$2I(x:x_e^*)=26.049$	12
f) $x(ijk.),x(..k\ell),x(ij.\ell)$		
RSU-effect | RST, RU, SU, TU	$2I(x_f^*:x_e^*)=9.752$	2
Interaction	$2I(x:x_f^*)=16.297$	10
g) $x(ijk.),x(ij.\ell),x(i.k\ell)$		
RTU-effect | RST, RU, SU, TU, RSU	$2I(x_g^*:x_f^*)=8.891$	4
Interaction	$2I(x:x_g^*)=7.406$	6
h) $x(ijk.),x(ij.\ell),x(i.k\ell),x(.jk\ell)$		
STU-effect | RST, RU, SU, TU, RSU, RTU	$2I(x_s^*:x_g^*)=4.543$	2
Third-order interaction	$2I(x:x_s^*)=2.863$	4

$x_h^* = x_s^*$

Table 8.3

Component due to	Information	D.F.
d) x(ijk.), x(i..l), x(.j.l)	$2I(x:x^*_d) = 28.754$	14
m) x(ijk.), x(ij.l)		
RSU-effect | RST, RU, SU	$2I(x^*_m:x^*_d) = 9.649$	2
Interaction	$2I(x:x^*_m) = 19.105$	12
f) x(ijk.), x(ij.l), x(..kl)		
TU-effect | RST, RU, SU, RSU	$2I(x^*_f:x^*_m) = 2.808$	2
Interaction	$2I(x:x^*_f) = 16.297$	10

Table 8.4

Component due to	Information	D.F.
b) x(ijk.), x(...l)	$2I(x:x^*_b) = 34.371$	17
c) x(ijk.), x(i..l)		
RU-effect | RST	$2I(x^*_c:x^*_b) = 5.303$	2
Interaction	$2I(x:x^*_c) = 29.068$	15
m) x(ijk.), x(ij.l)		
RSU-effect | RST, RU	$2I(x^*_m:x^*_c) = 9.963$	3
Interaction	$2I(x:x^*_m) = 19.105$	12
n) x(ijk.), x(ij.l), x(i.kl)		
RTU-effect | RST, RU, RSU	$2I(x^*_n:x^*_m) = 11.699$	6
Interaction	$2I(x:x^*_n) = 7.406$	6

$x^*_n = x^*_d$

Table 8.5

Computer Output x^{*}_{s}

RESIDUALS: R * S * T * U. FIRST 2 SUBSCRIPTS: 1 1

			1	2
1	OBSERVED	1	9.000000	6.000000
1	PREDICTED	2	10.241535	4.758002
1	RESIDUAL	3	-1.241535	1.241998
1	STANDARDIZE	4	-1.163043	1.391587
1	LOG RATIO	5	2.692764	1.927140
2	OBSERVED	6	8.000000	7.000000
2	PREDICTED	7	6.746210	8.254419
2	RESIDUAL	8	1.253790	-1.254419
2	STANDARDIZE	9	1.363679	-1.153870
2	LOG RATIO	10	2.276293	2.478061
3	OBSERVED	11	9.000000	6.000000
3	PREDICTED	12	9.012303	5.987520
3	RESIDUAL	13	-0.012303	0.012480
3	STANDARDIZE	14	-0.012295	0.012490
3	LOG RATIO	15	2.565903	2.156589

RESIDUALS: R * S * T * U. FIRST 2 SUBSCRIPTS: 1 2

			1	2
1	OBSERVED	1	14.000000	1.000000
1	PREDICTED	2	12.758455	2.241597
1	RESIDUAL	3	1.241545	-1.241597
1	STANDARDIZE	4	1.300076	-0.807367
1	LOG RATIO	5	2.913506	1.174679
2	OBSERVED	6	9.000000	6.000000
2	PREDICTED	7	10.253795	4.745572
2	RESIDUAL	8	-1.253795	1.254428
2	STANDARDIZE	9	-1.173809	1.407280
2	LOG RATIO	10	2.694960	1.924523
3	OBSERVED	11	12.000000	3.000000
3	PREDICTED	12	11.987688	3.012478
3	RESIDUAL	13	0.012312	-0.012478
3	STANDARDIZE	14	0.012308	-0.012452
3	LOG RATIO	15	2.851192	1.470075

RESIDUALS: R * S * T * U. FIRST 2 SUBSCRIPTS: 2 1

	1	2

1	OBSERVED	1	11.000000	4.000000
1	PREDICTED	2	11.685621	3.314257
1	RESIDUAL	3	-0.685621	0.685743
1	STANDARDIZE	4	-0.665103	0.752241
1	LOG RATIO	5	2.825670	1.565545
2	OBSERVED	6	14.000000	1.000000
2	PREDICTED	7	12.472881	2.527158
2	RESIDUAL	8	1.527119	-1.527158
2	STANDARDIZE	9	1.617001	-0.927095
2	LOG RATIO	10	2.890868	1.294407
3	OBSERVED	11	12.000000	3.000000
3	PREDICTED	12	12.841480	2.158594
3	RESIDUAL	13	-0.841480	0.841406
3	STANDARDIZE	14	-0.813287	0.987465
3	LOG RATIO	15	2.915992	1.136769

RESIDUALS: R * S * T * U. FIRST 2 SUBSCRIPTS: 2 2

			1	2
1	OBSERVED	1	9.000000	6.000000
1	PREDICTED	2	8.314383	6.685740
1	RESIDUAL	3	0.685617	-0.685740
1	STANDARDIZE	4	0.713138	-0.649305
1	LOG RATIO	5	2.485299	2.267289
2	OBSERVED	6	8.000000	7.000000
2	PREDICTED	7	9.527126	5.472840
2	RESIDUAL	8	-1.527126	1.527160
2	STANDARDIZE	9	-1.397613	1.722786
2	LOG RATIO	10	2.621455	2.067110
3	OBSERVED	11	11.000000	4.000000
3	PREDICTED	12	10.158509	4.841404
3	RESIDUAL	13	0.841491	-0.841404
3	STANDARDIZE	14	0.875411	-0.763642
3	LOG RATIO	15	2.685623	1.944516

RESIDUALS: R * S * T * U. FIRST 2 SUBSCRIPTS: 3 1

			1	2
1	OBSERVED	1	9.000000	6.000000
1	PREDICTED	2	8.882741	6.118252
1	RESIDUAL	3	0.117259	-0.118252
1	STANDARDIZE	4	0.118029	-0.117102
1	LOG RATIO	5	2.551422	2.178588
2	OBSERVED	6	13.000000	2.000000
2	PREDICTED	7	12.425370	2.574304
2	RESIDUAL	8	0.574630	-0.574304
2	STANDARDIZE	9	0.587705	-0.504864
2	LOG RATIO	10	2.887053	1.312891
3	OBSERVED	11	13.000000	2.000000

3	PREDICTED	12	13.691921	1.307411
3	RESIDUAL	13	-0.691921	0.692589
3	STANDARDIZE	14	-0.674136	0.850196
3	LOG RATIO	15	2.984118	0.635361

RESIDUALS: R * S * T * U. FIRST 2 SUBSCRIPTS: 3 2

			1	2
1	OBSERVED	1	11.000000	4.000000
1	PREDICTED	2	11.117260	3.881743
1	RESIDUAL	3	-0.117260	0.118257
1	STANDARDIZE	4	-0.116640	0.120037
1	LOG RATIO	5	2.775810	1.723597
2	OBSERVED	6	13.000000	2.000000
2	PREDICTED	7	13.574622	1.425695
2	RESIDUAL	8	-0.574622	0.574305
2	STANDARDIZE	9	-0.562285	0.676974
2	LOG RATIO	10	2.975513	0.721972
3	OBSERVED	11	15.000000	0.000005
3	PREDICTED	12	14.308073	0.692593
3	RESIDUAL	13	0.691927	-0.692588
3	STANDARDIZE	14	0.708390	-0.000059
3	LOG RATIO	15	3.028135	-0.000001

HYPOTHESIS 4 2I(X:X*) = 7.406 DEGREES OF FREEDOM =
6 PROBABILITY OF A LARGER VALUE = 0.284956

LOG(XSTAR/N/CELLS) = -2.382215

FIGURE 8.1

R i	S j	T k	U l	L	τ_1^R	τ_2^R	τ_1^S	τ_1^T	τ_2^T	τ_1^U	τ^{RS} 11	21	τ^{RT} 11	21	12	22	τ^{RU} 11	21	τ^{ST} 11	12	τ^{SU} 11	τ^{TU} 11	21	τ^{RST} 111	211	112	212	τ^{RSU} 111	211	τ^{RTU} 111	211	121	221	τ^{STU} 111	121	τ^{RSTU} 1111	2111	1121	2121
1	1	1	1	1	1		1	1		1	1		1				1		1		1	1		1				1		1				1		1			
1	2	1	1	1	1			1		1			1				1					1						.		1									
2	1	1	1	1		1	1	1		1		1		1				1	1		1	1			1				1		1			1			1		
2	2	1	1	1		1		1		1				1				1				1									1								
3	1	1	1	1			1	1		1									1		1	1												1					
3	2	1	1	1				1		1												1																	
1	1	2	1	1	1		1		1	1	1				1		1			1	1		1			1		1				1			1			1	
1	2	2	1	1	1				1	1					1		1						1									1							
2	1	2	1	1		1	1		1	1		:				1		1		1	1		1				1		1				1		1				1
2	2	2	1	1		1			1	1						1		1					1										1						
3	1	2	1	1			1		1	1										1	1		1												1				
3	2	2	1	1					1	1													1																
1	1	3	1	1	1		1			1	1						1				1							1											
1	2	3	1	1	1					1							1																						
2	1	3	1	1		1	1			1		1						1			1								1										
2	2	3	1	1		1				1								1																					
3	1	3	1	1			1			1											1																		
3	2	3	1	1						1																													
1	1	1	2	1	1		1	1			1		1						1					1															
1	2	1	2	1	1			1					1																										
2	1	1	2	1		1	1	1				1		1					1						1														
2	2	1	2	1		1		1						1																									
3	1	1	2	1			1	1											1																				
3	2	1	2	1				1																															
1	1	2	2	1	1		1		1		1				1					1						1													
1	2	2	2	1	1				1						1																								
2	1	2	2	1		1	1		1			1				1				1							1												
2	2	2	2	1		1			1							1																							
3	1	2	2	1			1		1											1																			
3	2	2	2	1					1																														
1	1	3	2	1	1		1				1																												
1	2	3	2	1	1																																		
2	1	3	2	1		1	1					1																											
2	2	3	2	1		1																																	
3	1	3	2	1			1																																
3	2	3	2	1																																			

FIGURE 8.2

	τ_1^R	τ_2^R	τ_1^S	τ_1^T	τ_2^T	τ_1^U	τ^{RS} 11	21	τ^{RT} 11	21	12	22	τ^{RU} 11	21	τ^{ST} 11	12	τ^{SU} 11	τ^{TU} 11	21	τ^{RST} 111	211	112	212	τ^{RSU} 111	211	τ^{RTU} 111	211	121	221	τ^{STU} 111	121	τ^{RSTU} 1111	2111	1121	2121
$n\pi$																																			
X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
X_a^+	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓			✓	✓				✓	✓	✓	✓												
X_b^+	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓				✓	✓	✓	✓												
X_c^+	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓	✓	✓												
X_d^+	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓												
X_e^+	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓												
X_f^+	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓										
X_g^+	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓						
$X_1^T = X_3^+$	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				
X_m^+	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓										
$X_g^T = X_n^+$	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓						

REFERENCES

[1] Bartlett, M.S. (1935), Contingency table interactions, J. Roy. Statist. Soc. Supplement 2, 248-252.

[2] Darroch, J.N. (1962), Interactions in multi-factor contingency tables, J.R. Statist. Soc. B. 24, 251-263.

[3] Ireland, C.T., Ku, H.H., and Kullback, S., (1969), Symmetry and marginal homogeneity of an r x r contingency table, J. Am. Statist. Assn. 64, 1323-1341.

[4] Ireland, C.T. and Kullback, S., (1968), Contingency tables with given marginals, Biometrika 55, 179-188.

[5] Ireland, C.T. and Kullback, S., (1968), Minimum discrimination information estimation, Biometrics 24, 707-713.

[6] Ku, H.H. and Kullback, S., (1968), Interaction in multi-dimensional contingency tables:an information theoretic approach, J. Res. Nat. Bur. Stand. B 72, 159-199.

[7] Ku, H.H., and Kullback, S., (1969), Approximating discrete probability distribution, I.E.E.E. Trans. on Information Theory, IT-15, 444-447.

[8] Ku, H.H., and Kullback, S. (1969), Analysis of multidimensional contingency tables:an information theoretic approach, I.S.I. Meeting, London, England, Sept. 1969, 156-158.

[9] Ku, H.H., Varner, R., and Kullback, S., (1968), Analysis of multi-dimensional contingency tables, Proceedings of the Fourteenth Conference on the Design of Experiments in Army Research Development and Testing, ARD-D Report 69-2, 141-180.

[10] Ku, H.H., Varner, R., and Kullback, S. (1971), On the analysis of multi-dimensional contingency tables, to appear in J. Am. Statist. Assn.

[11] Kullback, S., (1959), <u>Information Theory and Statistics</u>, Wiley, N.Y.; 1968, Dover Publications, Inc., N.Y.

[12] Kullback, S., (1968), Probability densities with given marginals, <u>Annals Math. Statist.</u> 39, 1236-1243.

[13] Kullback, S., (1970), Various applications of minimum discrimination information estimation, particularly to problems of contingency table analysis, <u>Proceedings of The Meeting on Information Measures</u>, University of Waterloo, Ontario, Canada.

[14] Kullback, S., (1971), Marginal homogeneity of multi-dimensional contingency tables, to appear in <u>Annals of Math. Statist</u>.

[15] Lewis, B.N., (1962), On the analysis of interaction in multi-dimensional contingency tables, <u>J.R. Statist. Soc. A</u> 125, 88-117.

[16] Roy, S.N. and Kastenbaum, M.A., (1956), On the hypothesis of no "interaction" in a multiway contingency table, <u>Annals Math. Statist</u>. 27, 749-57.

[17] Simpson, E.H., (1951), The interpretation of interaction in contingency tables, <u>J.R. Statist. Soc. B</u> 13, 238-241